Robust Registration of Narrow-Field-of-View Range Images

Stefan May [*] **Rainer Koch** [*] **Robert Scherlipp** [**]
Andreas Nüchter [***]

[*] *Faculty of Electrical Engineering and Fine Mechanics, Georg-Simon-Ohm University of Applied Sciences, Nuremberg, Germany (e-mail: firstname.lastname@ohm-university.eu).*

[**] *Healthcare Sector, Clinical Products Division, Siemens AG, Kemnath, Germany (e-mail: robert.scherlipp@siemens.com)*

[***] *Automation Group, School of Engineering and Science, Jacobs University Bremen gGmbH, Germany (e-mail: a.nuechter@jacobs-university.de)*

Abstract: This paper focuses on range image registration for robot localization and environment mapping. It extends the well-known Iterative Closest Point (ICP) algorithm in order to deal with erroneous measurements. The dealing with measurement errors originating from external lighting, occlusions or limitations in the measurement range is only rudimentary in literature. In this context we present a non-parametric extension to the ICP algorithm that is derived directly from measurement modalities of sensors in projective space. We show how aspects from reverse calibration can be embedded in search-tree-based approaches. Experiments demonstrate the applicability to range sensors like the Kinect device, Time-of-Flight cameras and 3D laser range finders. As a result the image registration becomes faster and more robust.

Keywords: range image registration, localization, robot vision

1. INTRODUCTION

With the objective of interacting with the real world, autonomous robotic research faces manifold problems at one go, e.g., fine motor skills, cognitive and planning abilities and in particular the robust perception of the environment. The scene has to be interpreted quickly especially in terms of motion estimation to avoid collisions and to perform manipulation tasks, needed, for instance, in rescue or medical robotics.

Range image registration is an important task in robotics for designing autonomous agents. The development of 3D sensors that provide high frame rates at low weight, energy consumption and costs pushes advances in robotic real-time perception. The algorithm design has to match sensor modalities, i.e., they have to be fast to meet real-time constraints and they need to deal with erroneous measurements dependent on the environment. There might occur measurement errors due to low or specular reflectivity, incidence of sunlight or when the distance between sensor and object is larger than the working range of the sensor. These issues result either in a wrong or a missing measurement point.

This paper compares derivatives of range image registration approaches based on the Iterative Closest Points (ICP) algorithm with respect to above-mentioned modalities. It further provides a formal classification of filters and an extension dedicated to narrow-field-of-view range sensors. The paper is organized as follows: Section 2 discusses

Fig. 1. Robot equipped with a Kinect device for 3D mapping purposes in indoor environments.

extensions to the ICP algorithm, all of which aiming at increasing robustness in range image registration. Section 3 classifies derivatives of ICP approaches and presents an extension and its intuitive derivation with respect to sensor modalities. The experiments are presented in section 4. Here, scene configurations are chosen in which range image registration is typically difficult. It is shown that the proposed extensions increase the robustness and convergence rate significantly for the selected scenes. Section 5 concludes the paper with an outlook on future work.

2. RELATED WORK

The *Iterative Closest Point* (ICP) algorithm, which was developed independently by Besl and McKay (1992), Chen and Medioni (1991) and Zhang (1992), is the most popular approach for range image registration. It aims at obtaining an accurate solution for the alignment of two point clouds – a reference, called model $M = \{\mathbf{m}_i | i = 1 \dots N_m\}$ and a scene denoted by $D = \{\mathbf{d}_i | i = 1 \dots N_d\}$ – by means of minimizing iteratively distances between point correspondences. The result can be expressed as a transformation matrix $\mathbf{T}$ consisting of a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$. In each iteration step the transformation minimizing an error function is computed. A point-to-plane error metric has been shown to be prior over a point-to-point error metric. Rusinkiewicz and Levoy (2001) provide a detailed analysis of efficient variants of the ICP approach, discussing the closed-form solutions, point-to-point vs. point-to-plane metrics, nearest neighbor assignment strategies and different rejection rules.

A generalized formulation of the ICP approach has been proposed by Segal et al. (2009). They showed that point-to-point and point-to-plane error metrics as well as a newly defined plane-to-plane error metric can be covered by a general formulation:

$$\mathbf{T} = \underset{\mathbf{T}}{\operatorname{argmin}} \sum_{i=1}^{N} \boldsymbol{d}_i^{(\mathbf{T})^\mathsf{T}} (\mathbf{C}_i^M + \mathbf{T}\mathbf{C}_i^D\mathbf{T}^\mathsf{T})^{-1} \boldsymbol{d}_i^{(\mathbf{T})}, \qquad (1)$$

where $\boldsymbol{d}_i^{(\mathbf{T})} = \boldsymbol{m}_i - \mathbf{T}\boldsymbol{d}_i$, $\mathbf{C}_i^M$ and $\mathbf{C}_i^D$ are covariance matrices associated with the model and the scene respectively. In dependence of these covariance matrices, one obtain one of the mentioned error metrics. The solution is found by iterating with a non-linear optimization step employing MLE.

The assignment of nearest neighbors is the most time consuming task in the iterative registration scheme. Measures of mainly two categories for reducing time consumption are widely used. The first one is based on search trees, which reduce complexity of search from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$. The second category comprise measures to find quickly close neighbors, but not necessarily the nearest ones. The price to be paid is the need for more iterations. But, if the speedup of one iteration is large enough, the total runtime can be significantly shorter. One early approach to be mentioned is called reverse calibration and makes use of projective geometry to find nearest neighbors along the line of sight of two intersecting point clouds (Blais et al., 1993; Neugebauer, 1997). Especially for GPU-based approaches, this search strategy is easy to implement and may use massive parallelism (Izadi et al., 2011).

Point correspondence rejection has a strong impact on the registration result. A straight forward implementation of the ICP approach assumes that the scene is completely covered by the model (Besl and McKay, 1992). In the case that the scene includes points which are not part of the model (from a non-overlapping or previously occluded area), wrong correspondences are assigned for these points which might distort the registration result (Fusiello et al., 2002). The simplest solution is the employment of a distance threshold. Corresponding tuples are rejected if their Euclidean distance exceeds this value. Several strategies are possible to determine suitable thresholds, e.g., a gradual decreasing threshold with respect to the iteration step. In general, these thresholds increase the registration performance on partially overlapping point clouds significantly, but are difficult to parameterize for not getting stuck in a local minimum.

Many extensions to the ICP approach have been published addressing the determination of valid point correspondences from overlapping parts. Chetverikov et al. (2002) proposed the *Trimmed ICP (TrICP)* approach. It employs a parameter representing the degree of overlap, i.e., the number of corresponding points N. Only the first N correspondence pairs ordered ascending by point-to-point distance are considered for estimating the optimal transformation. The disadvantage is that the degree of overlap needs to be estimated in advance, e.g., by the use of a second sensor. Prusak et al. (2007) employed, for instance, a spherical camera in order to obtain a coarse estimate for the degree of overlap.

Several approaches have been proposed to overcome the registration problem with unknown degree of overlap. Fusiello et al. employed the X84 rejection rule, which uses robust estimates for location and scale of a corrupted Gaussian distribution. It aims at estimating a suitable rejection threshold concerning the distance distribution between corresponding points. Niemann et al. (2003) proposed a rejection rule that considers multiple point assignments *(Picky ICP algorithm)*. If multiple points from the scene are assigned to the same corresponding model point, only the scene point with the nearest distance is accepted. The other pairs are rejected. Pajdla and Van Gool (1995) proposed the inclusion of a reciprocal rejection rule *(Iterative Closest Reciprocal Point algorithm - ICRP)*. For a corresponding point pair, which has been determined by searching the nearest neighbor of a scene point in the model, the search is reversed subsequently, i.e., for the corresponding model point the nearest neighbor in the scene is determined. This does not need to be the same scene point. The point correspondence is rejected if both scene points have a distance larger than a certain threshold. A disadvantage of the ICRP approach is the higher computational effort, since the nearest neighbor search, which is the most time consuming task, is performed twice as much as for all other approaches.

3. ICP EXECUTION CHAIN

Rejection rules included in the ICP approach can be classified in two categories: Pre-assignment filters and post-assignment filters. Pre-assignment filtering is performed before the time consuming nearest neighbor search, i.e., assignments of points are not known at this stage of computation. A related rejection rule can only employ attributes of the selected scene point.

In contrast, post-assignment filters employ the distance between assigned tuples. Both filter types can be included in the execution chain as summarized in Algorithm 1.

3.1 Projective Filtering

The following subsection explains the idea of employing the perspective projection of 3D measurements as a pre-

Algorithm 1. ICP iteration scheme

(1) Apply pre-assignment filter to select a subset D' from scene D.
(2) For each point in D' search nearest neighbor in M.
(3) Apply post-assignment filter to select a subset from the tuples $\{(\mathbf{m}_k, \mathbf{d}'_k)|k \in 1 \dots N'\}$ found in step 2.
(4) Calculate $\mathbf{T}$ for the tuples according to equation (1).
(5) Apply transformation to scene D.
(6) If termination criterion does not apply, go back to step 1.

assignment filter. The definition of this filter is directly based on projective geometry that underlies the measurement principle of a Kinect device or a Time-of-Flight (ToF) camera. The transformation which the sensor has undergone between two poses is to be determined from the sensor's data. Assuming, we would know the correct transformation, we could say which measurement point of the second data take could have been in line of sight from the first pose. In fact, this information is not known. But, the basic principle of the ICP algorithm is that the correct transformation is iteratively approached. Thus, the filter employs the estimated transformation during the iteration scheme resulting in a valid rejection rule. The main difference to reverse calibration is that the Euclidean distance is still used as a metric for finding nearest neighbors. Consequently, advantages of using perspective geometry is introduced for search-tree-based algorithms. It can be performed as follows. Homogeneous coordinates $\boldsymbol{\xi}_i$ of scene points are projected with the general perspective projection matrix $\mathbf{P}$ of the model to the image plane with

$$\mathbf{P}\boldsymbol{\xi}_i \rightarrow (u_i, v_i)^\mathsf{T}, \tag{2}$$

where u and v are resulting image coordinates from the perspective view of the model's pose. With a validity check, several aspects can be considered:

First, projections from the outer side of the visible frustum can be determined. The same effect can be achieved with a frustum check as previously published in (May et al., 2009). Being close to the correct solution, points from non-overlapping areas are removed without the need of parametrization, e.g., with a threshold.

Second, a projection to image coordinates in which an erroneous measurement was recorded before, can also be determined. This test relates, for instance, to specular reflections or the incidence of direct sunlight. The application of the projective filter is performed as follows:

$$D' = \{\boldsymbol{d}_i | z(u_i, v_i) > 0 \wedge z(u_i, v_i) < z_{max}\}, \tag{3}$$

where the coordinates (u_i, v_i) are obtained by projecting the query points with matrix $\mathbf{P}$, the function z relates the depth measurements and the value z_{max} denotes the maximum depth value of the model, i.e., the definition of the far frustum plane.

Finally, occlusions can be detected since at least two measurements would be projected to the same image coordinates. In computer graphics this rejection is called *z-Buffer test*. Only the tuple including the nearest scene point is considered, others are rejected.

$$D'' = \{\boldsymbol{d}'_i | \min z(u_i, v_i)\} \tag{4}$$

Summarized, each measurement is rejected, if the corresponding image projection could not have been in line of sight from the model's perspective.

3.2 Estimation of Surface Normals

For estimating surface normals, an area weighted method has been chosen, since it states a trade-off between fast computation and quality (Klasing et al., 2009):

$$\mathbf{n}_k = \frac{1}{N_q} \sum_{j=1}^{N_q} w_j \frac{([\mathbf{q}_{k,j} - \mathbf{m}_k] \times [\mathbf{q}_{k,j+1} - \mathbf{m}_k])}{|[\mathbf{q}_{k,j} - \mathbf{m}_k] \times [\mathbf{q}_{k,j+1} - \mathbf{m}_k]|}, \tag{5}$$

where N_q are the number of neighbors $\mathbf{q}_{k,j}$ around the model point. The weights w_j are computed as the area of the triangle spanned by these neighbors:

$$w_j = \frac{1}{2} |[\mathbf{q}_{k,j} - \mathbf{m}_k] \times [\mathbf{q}_{k,j+1} - \mathbf{m}_k]|. \tag{6}$$

Substituting equation (6) into equation (5) reduces the computation to the average of the cross products of adjacent triangles.

A more accurate solution can be implemented by fitting surface elements to the set of nearest neighbors of each query point. But due to the higher computational effort, we consider above mentioned approach to be prior w.r.t. the computational demand to a CPU-based implementation.

4. EXPERIMENTS AND RESULTS

The convergence of three different configurations were recorded within the iteration loop of each registration process. Initialization was set to identity. The projective filter has been embedded as a pre-assignment filter for point-to-plane and plane-to-plane error metrics. The benchmark has been performed against a gradual decreasing distance filter and reverse calibration.

4.1 Evaluation Measures

The convergence evaluation is performed on the basis of an axis-angle representation. The 4×4 matrices $\mathbf{T}_c$ to which a correct registration run needs to converge, have been determined manually by providing close estimates. Further, let $\mathbf{T}_{icp}$ be the estimated pose change determined for different registration runs. The angular error $e_{\Delta\theta}$ and the translational error $e_{\Delta\mathbf{t}}$ is then determined from the remaining pose difference computed by $\mathbf{T}_e = \mathbf{T}_c \mathbf{T}_{icp}^{-1}$.

4.2 Scene Configuration

In order to demonstrate that the proposed filter increases robustness the following three scene configurations were chosen:

(1) Compared to 3D laser range finders, the field of view is small by using real-time capable range sensors like ToF cameras or the Kinect device. When the rotational movement is fast, two data takes provide only a small partial overlap. Fig. 2 shows a scene in which a rotational movement of 26 ° between two data takes has been performed. Since the Kinect device has a horizontal field of view of 57 ° the overlap is only $\approx 54\%$.

Fig. 2. Scene 1 for evaluation: Left) Scene in line of sight during rotational movement. Center/Right) Two range images with low overlap were taken. Depicted is the perspective view of the Kinect device. It can be noticed that specular reflections avoid distance measurements when the emitted light hits the ground under an obtuse angle. Black areas were either not in line of sight or incapable of measurement due to reflections or external lighting.

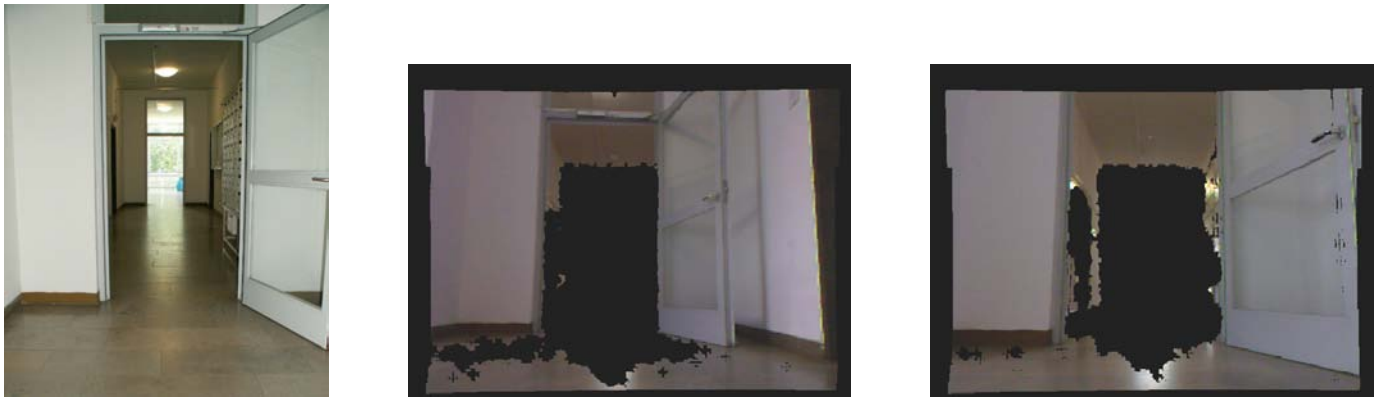

Fig. 3. Scene 2 for evaluation: Left) Scene in line of sight during translational movement. Center/Right) Perspective view of the Kinect device. It can clearly be seen that measurement points from objects out of the working range cannot be determined (here at the center of image). They will appear when the robot comes closer. Especially in this situation, the definition of a far frustum plane is meaningful. Additionally, occluded areas will be uncovered while approaching the open door.

Fig. 4. Scene 3 for evaluation: Left) Scene in line of sight for the data take with incidence of direct sunlight. Center/Right) Sunlight and low reflectivity avert the determination of distances with a Kinect device (cf. black area in the central part).

(2) Some occluded areas are disclosed from the sensor's perspective view when performing a translational movement. Fig. 3 shows a scene in which a robot was driven towards an open door. Here, a filter needs to determine areas that were not in line of sight from the pose of the first data take. Additionally, one can see that the limitation of the working range and specular reflections influence distance measurements.
(3) In some situations indoor robotics has to deal with external lighting. Fig. 4 shows a scene with incidence of sunlight through a window. Distance measurements cannot be determined from areas hit by direct sunlight.

4.3 Evaluation Results

The ICP approach used for the experiments employs the publicly available GNU Scientific library (GSL) and the Fast Library for Approximate Nearest Neighbors (FLANN). The runtime in each iteration step is not constant since the pre-filtering reduces the amount of nearest neighbor searches. Thus, for data sets with a small overlap the runtime decreases with each iteration step while approaching the minimum. Other optimization techniques like parallelization were not used. There is still potential in reducing runtime.

The entire scene covers 640×480 samples to be deducted by erroneous measurements, resulting in 307200 samples at most. The comparison has been performed on a subsampled point set selecting only every 25^{th} measurement point of the scene. For each registration task the runtime could be reduced. Table 1 compares the runtime exemplarily for the point-to-plane error metric. Here, the same amount of iterations were configured for both filters in order to compare the computational demand of the iteration loops. Additionally, a reduction in runtime can be obtained by the faster convergence, since less iterations would be necessary.

Table 1. Runtime comparison for point-to-plane metric with subsampled scene.

	Rev. Calib. (ms)	Distance (ms)	Projective (ms)
scene1	70	455	235
scene2	88	334	236
scene3	68	370	152

The runtime exclude the computation of surface normals. For the Kinect device, one has to be aware of the quality of distance data, when dealing with objects that are not directly in the close-up range. The number of nearest neighbors in equation (5) for estimating surface normals has to be adapted w.r.t. the image quality, which influences the overall runtime and convergence. The employment of both error metrics are sensitive to the quality of normals computation.

For all scenes either the rotational or the translational error is provided, since both dimensions reflected the same convergence results in our experiments. The registration results of the first scene are depicted on the left side of Figures 5 and 6. The scene of the underlying data set is mapped to the model with a significant rotation and a slight translation. Thus, only the rotational error is plotted. The reverse calibration approach did not converge for both error metrics. We assume that large discontinuities in depth between foreground and background are the reason for the poor performance.

Also, projective filtering introduces an error. The registration did not converge properly, since the starting pose of the scene is too far away from to the true value. The experiment had been modified in terms of filter activation. The filter has been applied only after a third of the iterations. The activation can clearly been seen in Figures 5 and 6, indicated by a faster drop off in the error metrics.

The registration results for the second scene are illustrated in the second column of both figures. The performed motion is dominantly translational, wherefore only the translational error is plotted. Here, also the reverse calibration showed poor performance. The reason for this behavior can also be explained by discontinuities in depth. Following a line of sight for the nearest neighbor search will assign incorrect neighbors at boundary areas, when previously occluded areas get uncovered. The Euclidean distance states a more stable metric in this situation. The application of projective filtering improved the registration results slightly.

For the scene with incidence of direct sunlight the results are differently. Reverse calibration performed best in terms of the necessary number of iterations as well as for the evaluated error metric. Either the point-to-plane and plane-to-plane metric were converging slowly towards the correct minimum for point assignments in Euclidean space. Nevertheless, the projective filter improves the convergence significantly.

4.4 Lessons Learned

The performance of the ICP approach depends on several factors. Since many extensions to the ICP approach have been proposed, a benchmark can only cover a smaller part of the combinatorial possibilities w.r.t. error metrics, assignment or filtering techniques or subsampling strategies. Most of the combinatorial complexity is added with the chosen environment and the performed sensor motion. With a benchmark covering only a smaller part of these aspects, one can produce nearly any desired results. Therefore, we disclaimed to design the evaluation w.r.t. to the proposed method.

5. CONCLUSIONS AND FUTURE WORKS

This paper presented an extension for range image registration based on the well-known Iterative Closest Point algorithm in order to deal with erroneous measurements. A classification of extensions to the ICP algorithm has been described. Depending on the integration into the execution chain, rejection rules can be classified as pre-assignment or post-assignment filters. In this context we presented an extension that is derived directly from measurement modalities of sensors in projective space. The intuitive formulation of the problem resulted in a pre-assignment filter that increased the convergence rate for k-d tree based assignment techniques in several situations. We showed that the principle idea of reverse calibration can be embedded in these algorithms.

Future work will focus on the impact of sensor motion to registration results. The experimental results that we

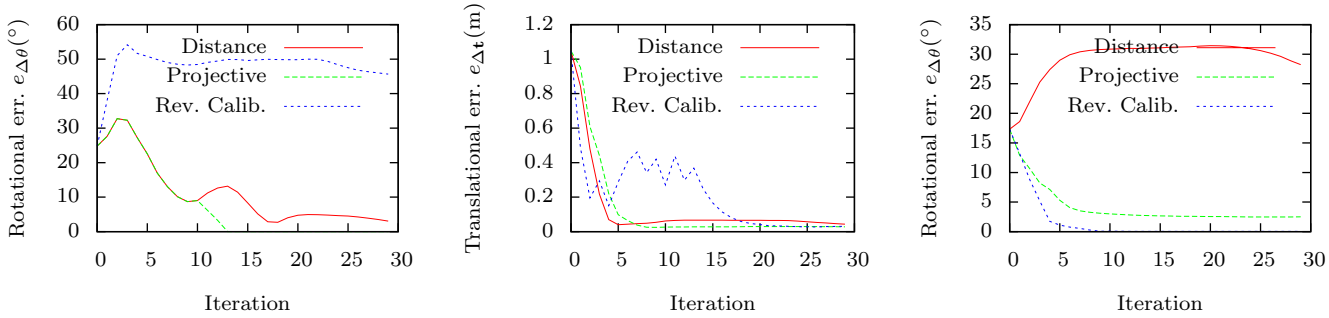

Fig. 5. Convergence for point-to-plane metric for scenes 1 to 3 from left to right.

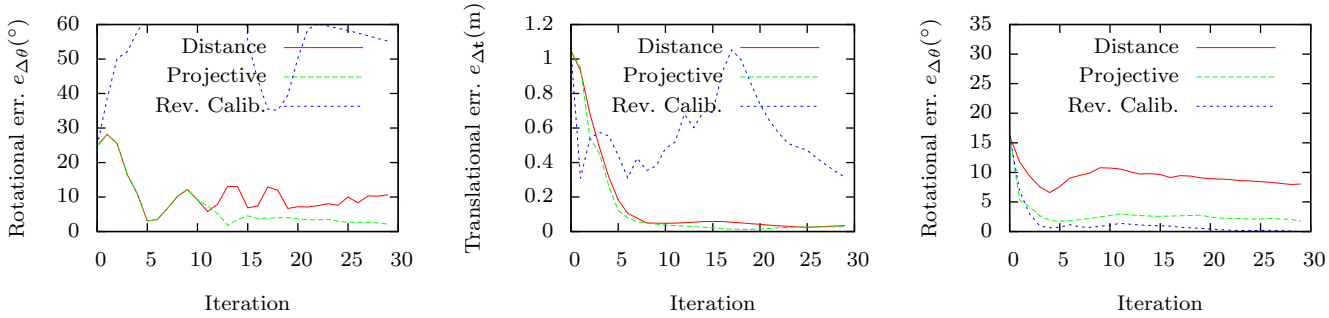

Fig. 6. Convergence for plane-to-plane metric for scenes 1 to 3 from left to right.

presented in this paper motivate for a detailed analysis in order to rate expected performance of ICP configurations w.r.t. the performed sensor motion.

REFERENCES

Besl, P. and McKay, N. (1992). A method for Registration of 3–D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239 – 256.

Blais, G., Levine, M.D., and Levine, M.D. (1993). Registering Multiview Range Data to Create 3D Computer Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8), 820–824.

Chen, Y. and Medioni, G. (1991). Object modeling by registration of multiple range images. In *In Proceedings of the IEEE Conference on Robotics and Automation (ICRA)*, 2724–2729. Sacramento, CA, USA.

Chetverikov, D., Svirko, D., Stepanov, D., and Krsek, P. (2002). The trimmed iterative closest point algorithm. In *Proceedings of the 16th International conference on pattern recognition (ICPR)*, volume 3, 545–548. Quebec.

Fusiello, A., Castellani, U., Ronchetti, L., and Murino, V. (2002). Model acquisition by registration of multiple acoustic range views. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen (eds.), *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, number 2351 in Lecture Notes in Computer Science, 805–819. Springer, Copenhagen, Denmark.

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. (2011). KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, 559–568. ACM, New York, NY, USA.

Klasing, K., Althoff, D., Wollherr, D., and Buss, M. (2009). Comparison of surface normal estimation methods for range sensing applications. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 3206–3211. Kobe, Japan.

May, S., Droeschel, D., Holz, D., Fuchs, S., Malis, E., Nuechter, A., and Hertzberg, J. (2009). Three-dimensional mapping with time-of-flight cameras. *Journal of Field Robotics (JFR), Special Issue on Three-dimensional Mapping*, 26(11–12), 934–965.

Neugebauer, P.J. (1997). Geometrical cloning of 3D objects via simultaneous registration of multiple range images. In *Proceedings of the 1997 International Conference on Shape Modeling and Applications (SMA '97)*, 130–139. IEEE Computer Society, Washington, DC, USA.

Niemann, H., Zinßer, T., and Schmidt, J. (2003). A refined ICP algorithm for robust 3-D correspondence estimation. In *Proceedings of the IEEE International Conference on Image Processing*. Barcelona.

Pajdla, T. and Van Gool, L. (1995). Matching of 3-D curves using semi-differential invariants. In *Proceedings of the Fifth International Conference on Computer Vision (ICCV)*, 390–395. Boston, MA, USA.

Prusak, A., Melnychuk, O., Roth, H., Schiller, I., and Koch, R. (2007). Pose Estimation and Map Building with a PMD-Camera for Robot Navigation. In *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D)*. Heidelberg, Germany.

Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the ICP algorithm. In *Proceedings of the Third International Conference on 3D Digital Imaging and Modellling (3DIM)*. Quebec City, Canada.

Segal, A.V., Haehnel, D., and Thrun, S. (2009). Generalized-ICP. In *Robotics: Science and Systems*.

Zhang, Z. (1992). Iterative point matching for registration of free-form curves. Technical Report RR-1658, INRIA Sophia Antipolis, Valbonne Cedex, France.